



Pay-for-Performance and Employee Mental Health

Large Sample Evidence Using Employee Prescription Drug Usage

Dahl, Michael S.; Pierce, Lamar

Published in:
Academy of Management Discoveries

DOI (link to publication from Publisher):
[10.5465/amd.2018.0007](https://doi.org/10.5465/amd.2018.0007)

Publication date:
2020

Document Version
Accepted author manuscript, peer reviewed version

[Link to publication from Aalborg University](#)

Citation for published version (APA):
Dahl, M. S., & Pierce, L. (2020). Pay-for-Performance and Employee Mental Health: Large Sample Evidence Using Employee Prescription Drug Usage. *Academy of Management Discoveries*, 6(1).
<https://doi.org/10.5465/amd.2018.0007>

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal -

Take down policy

If you believe that this document breaches copyright please contact us at vbn@aub.aau.dk providing details, and we will remove access to the work immediately and investigate your claim.

Pay-for-Performance and Employee Mental Health: Large Sample Evidence Using Employee Prescription Drug Usage

Michael S. Dahl
Aarhus University
Department of Management
Building 2628, 315
8210 Aarhus V Denmark
msd@mgmt.au.dk

Lamar Pierce
Washington University in St. Louis
Olin Business School
One Brookings Drive Box 1156
St Louis, MO 63130 USA
pierce@wustl.edu

Forthcoming in *Academy of Management Discoveries*

This project was funded by the Independent Research Council | Social Science (Case no. 0602-02540B). We thank Søren Leth-Sørensen at Statistics Denmark for valuable assistance. Anders Frederiksen, Andrew Knight, and Ian Larkin provided valuable comments.

ABSTRACT: This paper provides evidence linking pay-for-performance (P4P) adoption by employers to long-term and serious mental health problems in employees. Matching survey-based data on P4P adoption by 1,309 Danish firms with wage, demographic, and medical prescription data of 318,717 full-time employees, we find a four to six percent increase in the usage of anti-depressant and anti-anxiety medication after firms adopt P4P. This change is strongest in low-performing and older workers. We also find that workers select in and out of P4P firms based on mental health considerations, which implies that mental health effects influence turnover. We similarly show that low performers are more likely to leave following P4P adoption. Finally, we show sizable but imprecise response differences from female and male employees to the mental health threat of performance-based pay. Women with latent or potential mental health concerns appear to leave firms after P4P adoption, while men do not. Although we cannot claim a causal relationship, collectively our results suggest a model where performance-based pay forces many employees to choose between leaving or else depression and anxiety. Our study expands existing work by showing that the mental health costs of performance-based pay can be severe enough to necessitate pharmaceutical treatment.

KEYWORDS: Incentives, Compensation, Mental Health, Stress, Depression

Performance-based pay is widely used by firms to both motivate employee effort and attract the best talent. Theoretical models in economics (Hölmstrom, 1979; Jensen & Murphy, 1990), psychology (Gerhart & Rynes, 2003; Vroom, 1964), and management (Gomez-Mejia & Welbourne, 1988; Nyberg, Pieper, & Trevor, 2016) argue that well-designed pay-for-performance (P4P) can improve worker performance by linking effort with financial rewards. These theoretical predictions are supported by evidence across fields (Prendergast, 1999; Rynes, Gerhart, & Parks, 2005), in settings such as automotive service (Lazear, 2000), agriculture (Bandiera, Barankay, & Rasul, 2005), trucking (Burks, Carpenter, Goette, & Rustichini, 2009), professional services (Hitt, Bierman, Shimizu, & Kochhar, 2001), and software sales (Larkin, 2014). Just as importantly, firms can benefit as the best workers may be attracted to superior pay under P4P, while the lowest performers instead seek hourly or salaried pay (Zenger, 1994; Cadsby, Song, & Tapon, 2007; Dawling & Falk, 2011; Trevor, Reilly, & Gerhart, 2012; Shaw, 2015). In contrast, P4P is limited by worker preferences for pay certainty (Cadsby et al., 2007; Dawling & Falk, 2011; Prendergast, 1999), noisy performance measures (Baker, 2000), multitasking problems (Holmstrom & Milgrom, 1991), gaming (Frank & Obloj, 2014; Larkin, 2014), motivational crowding out (Frey, 1997; Ryan & Deci, 2000; Frey & Jergen, 2001; Benabou & Tirole, 2003), and social comparison costs and envy through pay disparity (Nickerson & Zenger, 2008; Larkin, Pierce, & Gino, 2012; Edelman & Larkin, 2014; Feldman, Gartenberg, & Wulf, 2018; Gartenberg & Wulf, 2017; 2018; Obloj & Zenger, 2017).

Recent work using laboratory experiments and surveys raise the spectre of another major cost of performance-based pay—mental health problems (Allen, Bender, & Theodossiou, 2017; Cadsby et al., 2016; Davis, 2016). Despite this important study, we know little about how serious and persistent these mental health costs might be, and whether they impact long-term employment relationships, fundamentally hurt employee wellness, and alter career paths. Understanding whether

these short-term effects extend to serious and long-term anxiety and depression is crucial because of severe economic and social costs of mental health that include medical costs, presenteeism, absenteeism, suicide, and spillover effects to friends and family (Greenberg, Fournier, Sisisky, Pike, & Kessler, 2015). Such understanding can also help us build more comprehensive models of employee response to compensation policy changes. The persistence and severity of effects from performance-based pay introduction are best answered through panel data that track long-term changes in firm compensation policy, employment, and mental health.

We provide new evidence for answering these questions, showing large-scale medical evidence of the relationship between pay-for-performance and mental health using individual data on wages, employment, and prescription drug usage. By matching these data with information on the implementation date of pay-for-performance at 1,309 large employers, we can observe how the introduction of P4P correlates with the use of benzodiazepine for anxiety and insomnia and selective serotonin reuptake inhibitors (SSRI) for anxiety and depression. In addition, we can observe whether P4P implementation is associated with selective turnover that suggests workers sorting in and out of jobs based on the mental health impact of compensation systems.

Using data from 318,717 full-time workers, we find evidence that pay-for-performance is indeed associated with mental health problems. Worker fixed-effect models that control for time-invariant differences show that when firms implement P4P, existing workers increase stress medication usage by 5.7% over base rate. We note that this increase represents slightly over 2,000 additional prescription-years in our data. Models using firm fixed effects show smaller increases of 4.4% that indicate workers with either untreated mental health issues or risk aversion disproportionately leave firms following P4P adoption and are replaced by those without such issues. This suggests that mental health problems not only increase following P4P adoption, but also motivate job changes as a result. We note that unlike stylized experimental studies (e.g., Eriksson &

Villeval, 2008; Cadsby, Song, & Tapon, 2016), we cannot strongly claim causality; the adoption of P4P by employers is endogenous, as is the employment choice of workers. The unique strength of our study is its panel of rich individual wage and objective medical data for all employees that can show persistent and serious mental health costs. To the best of our knowledge, no prior work can demonstrate these costs.

We show several additional effects of P4P on workers. First, although workers appear on average to enjoy slightly higher average earnings under P4P, those with both the best pre-adoption and post-adoption wage trajectories appear to suffer no effect. This indicates that the mental health increases from P4P adoption are associated with lower performers who earn less under performance-based pay. Second, we find evidence that the endogenously sorting based on unobservable mental health problems in our data occurs with women. Although women who remain at firms after P4P adoption are equally likely as their male counterparts to suffer mental health problems, those with the highest propensity for stress appear to leave firms following P4P adoption, in sharp contrast to men. Third, our results appear to be primarily driven by employees over the age 50, who tend to have limited job mobility. These subsample results suggest the need for a more comprehensive model and additional empirical work that can help answer precisely how employees respond to P4P adoption and the moderators associated with these responses.

EMPLOYEE RESPONSES TO PAY-FOR-PERFORMANCE

Existing theory and evidence suggest that when an employer adopts pay-for-performance, the employee may be affected in several ways. P4P might directly increase anxiety and depression through multiple mechanisms. First, if the employee inherently dislikes risk and uncertainty, as economics (Prendergast, 1999), psychology (Rynes et al. 2005), and management (Larkin et al. 2012) argue, the increased pay uncertainty in itself could generate mental health concerns. Second, the

greater pay variance across workers with similar jobs that is generated by P4P might evoke social comparison (Festinger, 1954; Larkin et al., 2012), envy (Nickerson & Zenger, 2008), and perceptions of inequity and unfairness (Fehr & Schmidt, 1999; Kim, Weber, Leung, & Moramoto, 2009) with coworkers and other peers. Third, P4P might induce stress if it increases employees' propensity to think of time as money (Pfeffer & Carney, 2018). Each of these mechanisms could directly increase anxiety or depression in all but the top-performing employees under the new P4P scheme.

In addition, mental health concerns might increase because of the cultural change commonly associated with P4P adoption and other incentives (Gneezy, Meier, & Ray-Biel, 2011). P4P systems can motivate competitive behaviors and disincentivize prosociality, particularly in systems that rely on individual performance (Chan, Li, & Pierce, 2014a, 2014b), tournament-based (relative) pay (Garcia, Tor, & Gonzalez, 2006; Garcia & Tor, 2007), or the peer-based division of rewards (Pierce, Wang, & Zhang, 2019). Tournament-based P4P is even thought to generate sabotage behaviors that are toxic for organizational culture (Drago & Garvey, 1998; Lazear, 1999; Charness, Masclet, & Villeval, 2013), even when performance is defined at the team level (Gürtler, 2008). Given the evidence that prosocial and cooperative cultures tend to foster better wellness (Bolino & Grant, 2016; Knight, Menges, & Bruch, 2017), such cultural change could increase mental health problems.

Pay-for-performance adoption might also shift employee mental health through the direct effect it has on their average wage income. P4P systems inherently raise the income of high performers while reducing pay for low performers. Changes in income can have strong effects on mental health, particularly for those living under debt or tight financial constraints (Bridges & Disney, 2010; Sweet, Nandi, Adam, & McDade, 2013; World Health Organization, 2014). An increase in income for high-performers might relax budget constraints, reducing stress around the affordability of basic needs such as housing, food, and the needs of children and other dependents. Furthermore, it might allow consumption of goods or services that improve mental health either by

freeing up time (e.g., outsourcing lawn care, cleaning) or providing relaxation or entertainment (e.g., vacation, fitness activities, entertainment).

In contrast, a pay reduction for low performers under P4P could have marked effects in increasing mental health problems for many of the same reasons raised above. In addition to these increased stressors, income reductions can also evoke the disutility of loss aversion (Kahneman, Knetsch, & Thaler, 1991). People are particularly hurt by losses such as income reduction, which can become manifested through anxiety and depression. Prospect theory suggests that if P4P generates a mean-preserving income spread, the mental costs to those who lose income will outweigh the gains of those who gain it. This suggests that while income change will have a negative relationship with mental health problems, the net outcome on average from increased pay variance might be worse mental health.¹

In addition, P4P adoption might hurt mental health through the physical health problems it generates. DeVaro & Heywood (2016), for example, find higher physical ailments and absence among workers employed at firms in the United Kingdom, consistent with a survey-based literature linking P4P to greater risk-taking and exertion that can generate accident and injury (Foster & Rosenzweig, 1994; Freeman & Kleiner, 2005; Bockerman, Johansson, & Kauhanen, 2012; Bender & Theodossiou, 2014; Artz & Heywood, 2015). Such physical health problems might generate or exacerbate mental health problems directly. Additionally, attempts to cope with mental health from P4P through drug and alcohol usage might backfire, generating *both* physical and additional mental health problems. Recent work by Arz, Green, and Heywood (2018) finds that P4P adoption increases such coping, which is known to generate a broad spectrum of additional health issues.

¹ We note that the removal of P4P would produce similar effects, as the cost to high earners losing income would outweigh the gains to low earners. Our data do not allow us to test this.

Finally, we note that in a labor market where some mobility exists, some employees will leave companies that adopt pay-for-performance as they re-sort into their preferred compensation systems. Economics and other fields argue that one of the most important implications of P4P is that it attracts the highest performers while repelling the worst, due to both rational expectations about pay outcomes (Lazear, 1986, 2000; Booth & Frank, 1999; Eriksson & Villeval, 2008; Cornelissen, Heywood, & Jirjahn, 2011; Shaw, 2015) and risk preferences (Dohmen & Falk 2006; Cadsby et al., 2016). Although overconfidence (Zenger, 1994; Larkin & Leider, 2012) or pay preferences (Hamilton, Nickerson, & Owan, 2003) might dull this effect, employees are more likely to leave a job after it becomes less remunerative or enjoyable.

This sorting mechanism is typically modeled around individual differences in performance (Lazear, 2000; Shaw, 2015). The introduction of P4P should raise the wages of the highest performers, motivating them to stay while also attracting similar high-performers from outside the firm. Similarly, low-performers would be more motivated to leave or avoid employment at the firm with a new P4P system. So even in the absence of mental health effects, P4P should induce performance-based sorting associated with wage changes.

If P4P induces additional psychological costs such as mental health problems, this sorting on performance should intensify, as increased wage dispersion changes not only the financial choice to leave, but also the emotional and psychological costs of staying. If pay comparisons with both others and prior year wages are negative, then mental health problems will emerge or intensify, which might in turn further motivate departure. But the effect of P4P on turnover need not only operate through wage changes. Other individual differences in preferences for or tolerance of a P4P system might drive turnover, independent of wages. Employees who suffer psychological costs from risk or loss aversion (Cadsby et al., 2016), organizational change (Begley & Czajka, 1993), or envy (Larkin, Pierce, & Gino, 2012) might leave to reduce these costs.

Our study uniquely examines severe cases through the use of medication data. Collectively, prior work predicts that P4P adoption should on average increase mental health problems, particularly for those whose wage income decreases. They also suggest selective turnover following P4P adoption as workers sort in and out of the firm based on performance (and the pay associated with it) and likely mental health under P4P because of risk preferences and tolerance for organizational change and competition. Sorting after P4P adoption would also be constrained by outside labor market mobility. Although we cannot directly measure these mechanisms behind mental health and turnover, we will attempt to address them by examining mental health changes in subsamples of workers consistent with these mechanisms. Because we are limited by statistical power only to subsamples large enough to plausibly identify true effects, we will examine employees by age, gender, and wage trajectory.

DATA AND METHOD

Data

We study pay-for-performance and mental health using three datasets linked through individual social security numbers in Denmark. Denmark is well-suited for this study due to both the characteristics of the labor market and data availability. The labor market is one of the most flexible and mobile in Europe with annual job-to-job mobility rates on par with the United States (Frederiksen & Westergaard-Nielsen, 2007; Dahl & Sorenson, 2010). If the introduction of P4P has negative effects on employees, this labor market flexibility allows them to easily sort into other firms. Any effects of P4P here could thus be seen as a conservative estimate that would be higher in less flexible and mobile labor markets.

The complete Danish social security system enables accurate matching of individual- and firm-level databases. Our first dataset, the Integrated Database for Labor Market Research (IDA),

contains annual demographic and employment information from the Danish government for all individuals in Denmark from 1980-2006 (see Eriksson and Lausten (2000) for a study linking P4P to firm performance with similar Danish registry data). These data identify family status, age, education level, gender, employer, and annual wages. We match these data with the second dataset, the Danish Register of Medicinal Product Statistics (RMPS), maintained by the Danish Medicines Agency, which includes all prescriptions for the entire Danish population from 1995 to 2006. This combination of demographic and medical data has previously been used by Dahl, Nielsen and Mojtabai (2010), Dahl (2011), and Pierce, Dahl, and Nielsen (2013). An advantage of studying employment and medical outcomes in Denmark is that employment changes will not affect medical care access. Although patients must pay for medication, amounts decrease substantially with the number of prescriptions. Low-income patients receive public support, and visits to general physicians are free of charge. We primarily focus on two different types of medication classified by ATC-classes, following Dahl, Nielsen and Mojtabai (2010) and Dahl (2011). Insomnia and anxiety are treated with benzodiazepine-related medications (ATC: N05CF and N05BA). Anxiety and depression are treated with selective serotonin re-uptake inhibitors (SSRI) (ATC: N06AB). We note that these data represent the complete medication usage of the employees in our dataset.

Our third dataset is survey data from the second wave of the DISKO survey (DISKO2) conducted by Statistics Denmark in Winter 2000/2001. The survey, used by several previous studies (Laursen & Foss, 2003; Foss & Laursen, 2005; Dahl, 2011), contains information on innovation, human resource practices, organizational change, and the use of P4P from the largest Danish firms. The survey is fourteen pages long and contains 50 different questions on these issues. We use information from two questions presented in Figure 1, where P4P is one of ten different HRM practices. For each of these ten HRM practices, Question 8 asks “*Does the firm make use of some of the following ways of planning the work and paying the employees?*” Question 9 requests additional information

on these ten practices *When were these measures introduced and how many employees are included (percentage)?*’.

The category of interest is listed last in these questions as “Performance related pay (not piece-rate pay)”.²

-----INSERT FIGURE 1 HERE-----

The DISKO2 survey was sent to the first wave respondents (DISKO1 was from 1996-1997) and all other firms in Denmark with more than 25 employees (see Dahl (2011) for further information). In total, 6,975 firms received the DISKO2 questionnaire. 2,162 firms (about 31%) responded to at least part of the survey. When we focus only on respondents that gave definitive answers for the two questions relevant to this study (we excluded those answering “Don’t know”), the number of firms drop to 1,309. These firms come from all private sector industries. The largest 2-digit SIC industries in the sample are construction and wholesale trading, each covering 17 percent of the firms. Nine percent of the construction firms and 21 percent of the wholesale firms are adopting P4P in the period (i.e. after 1995). The third largest is business services (law, advertising, etc.) covering 8 percent of the firms, where 21 percent adopt P4P in this period. At the more aggregate level, 35 percent of the firms are in the manufacturing sector, the remaining are construction, trade and other services.

These survey data have several weaknesses. First, we cannot observe which specific employees received P4P within the firm, such that any effect that we estimate will be averaged across all employees. We will address this by later examining employees based on their wage changes. Second, we can only observe if firms historically used P4P during our sample period and then withdrew it if they answered the same question on the DISKO1 survey in 1996. Only nine firms indicated to have dropped P4P in the period between surveys, but because the survey does not

² “Piece-rate” is translated from the Danish “akkord”, which refers to production quotas, and does not imply performance-based pay as it might in English.

indicate in which year this occurred, we cannot exploit this change. This second weakness is important because some of our theoretical mechanisms, such as endogenous labor market sorting and loss aversion, would also apply to the removal of P4P. We note that other mechanisms, such as risk aversion, social comparison, and cultural changes, would not apply to such a removal.

An additional weakness is that we cannot differentiate between classes of pay-for-performance. P4P schemes differ based not only on level (individual vs. team) as well as structure (e.g., fixed bonus, straight commission, or non-linear scheme), with complex implications of these structures for the economic and psychological responses of workers (Gerhart & Rynes, 2003; Larkin, Pierce, & Gino, 2012; Nyberg, Pieper, & Trevor, 2016). We note that although the survey item explicitly excludes “piece-rate pay”, this is only one of many classes of P4P.

Finally, since the survey ended in 2001, we cannot observe P4P adoption or cancellation between 2002 and 2006. We note that these limitations bias against finding a relationship between P4P and mental health concerns because they falsely code many “treated” employees as “untreated.” One way to address this limitation would be to simply limit our individual-level data to 1996-2001, but this approach suffers several key problems.³ First, because both mental health problems and the choice to treat them may be delayed, a shorter sample would miss many of the later effects of P4P adoption, particularly in firms that adopt in later years. As we demonstrate in our leads and lags model later, effect sizes increase with each year following P4P adoption, as workers eventually seek medical treatment. Second, ending the data in 2001 decreases the number of “treated” worker-years by 35.7%. Despite the many observations in our sample, our statistical power is already limited by a low base rate (4.5%), and small effect size (0.29%), and a high intraclass correlation (0.44). Even

³ See the Appendix for coefficient estimates from worker fixed effect models using samples with alternative ending years ranging from 2001 to 2005.

with the our full sample of 1996-2006, our statistical power is only 0.456 to identify our primary effect size of 5.6%.⁴

We link the RMPS and IDA individual data with the survey data via unique firm ID. Our combined data therefore covers 1,159,417 person-years—318,717 unique full-time employees at 1,309 firms between 1995 and 2006. Each worker-year observation therefore identifies wages, demographics, medications, employer, and employer responses to DISKO survey questions. We are restricting the sample to full-time workers between the ages of 18 and 65. Individuals close to the age of 65 will have access to early retirement, which might influence their response to the introduction of P4P differently.

Of these 1,309 firms, 445 indicate using some level of P4P in the survey (Question 8). Our empirical model with firm fixed effects will link mental health to within-firm changes in P4P from the 242 firms that adopted P4P between 1996 and 2001 (Question 9). The average firm in our sample has 80 employees in a given year. Figure 2 presents the distribution of P4P adoption year from Question 9, with vertical lines indicating the time range of our data. We note that our models will identify treatment effects of P4P only off those firms who adopt during this period, since other firms' exclusive use of P4P or not during our time period will be absorbed by firm fixed effects.

-----INSERT FIGURE 2 HERE-----

Dependent variables. Our primary dependent variables are indicators from the RMPS dataset that an employee uses a given class of medication in a particular year. The first class is *benzo*, which indicates that the individual used benzodiazapines in that year. The second is *ssri*, which indicates the use of SSRI medication. The third, *stress*, indicates the use of at least one of these drug categories. Like Dahl (2011) and Pierce et al. (2013), we note a key weakness with measuring mental

⁴ Power was calculated through 10,000 simulated datasets that matched our data based on observations, number of firms, number of workers, number of years, stress base rate, an effect size of 0.0029, and a worker intracluster correlation of .474.

health through prescription drug usage that generates measurement error but should not bias our model estimates. We are observing treatment, not problems, so many employees may have untreated mental health issues. A recent study of American men found that only about 30% of those with anxiety or depression took medication as treatment (Blumberg, Clarke, & Blackwell, 2015). Our observation of treatment is conditioned on employees visiting their GP or leaving a hospital with a prescription. We note that this under-measurement should bias against us finding results. We also note that alternative measures of mental health concerns, typically surveys, suffer arguably larger problems in self-response bias. The ideal of mandatory clinical diagnosis is clearly not feasible in this type of study.

Independent variables. Our key independent variable is a dummy indicating that the firm reported using pay-for-performance in a given year. We construct the pfp indicator based on the reported date of PFP adoption in Question 9 of the DISKO2 survey data. If, for example, a firm reports that it adopted P4P in 1998, then all employees of that firm will be coded zero for 1996 and 1997 and one for 1998 through 2006. Firms that adopted P4P before 1997 are always coded one. Firms that answered “no” to P4P adoption in Question 8 (and therefore did not report a date for Question 9) are always coded zero. This variable will serve as the “treatment” variable in our difference-in-differences specification described below. Alternatively, we use $pfppercent$ as the logged percentage of compensation tied to performance in a given year, as indicated by the firm. This variable allows the relative frequency of P4P usage within the firm to differentially affect mental health.

Control variables. We include key demographic control variables that might also influence mental health concerns and are potentially correlated with adoption of P4P via demographic differences between firms. Family variables include the number of children (under 18) and marriage and domestic partner status. Demographics include gender, a cubic age function, and education

level. All specifications within the paper also control for time trends with year dummies. We note that any time-invariant (or nearly time-invariant) demographic control will be dropped from our primary worker fixed effect models, but are important in our secondary firm fixed effects models.

Table 1 presents descriptive statistics and correlations for the 1,159,417 employee-year observations in our sample. Our key dependent variable, *stress*, indicates that 5.2% of employees are using either (or both) benzodiazepines (3.9%) or SSRIs (1.8%) in a given year, with slightly over half (52.4%) working a firm with a P4P system. The average firm reports using P4P for 14.3% of its employees, with 37.4% of workers under P4P in those firms with P4P plans. It's important to note that the slightly negative raw correlation between *pfp* and *stress* is not evidence against an effect, since it ignores (among other things) fixed differences in the firms and employees that adopt P4P. Our primary worker fixed effect models will observe changes in medication use within workers after P4P is adopted.

-----INSERT TABLE 1 HERE-----

Empirical Approach

Given the panel structure of our data, we adopt a difference-in-differences strategy (DiD) that compares the change in employee prescription drug usage as firms adopt pay-for-performance, relative to companies that do not. The baseline model uses a linear probability model with worker fixed effects, represented by

$$Y_{ijt} = \alpha_{ij} + \beta_1 * PFP_{jt} + \beta_2 * FirmSize_{jt} + \beta_3 * Demographics_{it} + \lambda_t + \varepsilon_{ijt}$$

The worker fixed effect α_{ij} is defined for the given employment relationship between worker i and firm j . PFP_{jt} is a dummy indicating that firm j indicated it used pay-for-performance in year t , $FirmSize_{jt}$ is total employment, and $Demographics$ indicate for worker i the number of children under

18, gender, a cubic age function, education level (in months), and whether the individual is married or in a domestic partnership. λ_t is a set of year dummies for both firms that use or do not use P4P. Because of the worker fixed effects, the parameter of interest β_i indicates the change in probability that a given employee who stays at the firm uses medication after the firm adopts pay-for-performance. These worker fixed effect models will only estimate coefficients on individuals who worked at the firm both before and after P4P adoption and thereby represent an average treatment effect on existing employees.

We note that although a logit model might also be used to predict our binary dependent variable, this approach is not appropriate in our case because our DiD model with staggered adoption requires either individual or firm fixed effects. Conditional logit models with individual fixed effects produce severely biased parameter estimates because the low number of observations for each worker suffers from the “incidental parameters problem” (Lancaster, 2000; Katz, 2001). A conditional firm fixed effects logit model is impossible in Stata software due to numerical overload problems.⁵ Given this, a linear probability model is most appropriate. To account for violations of the standard error assumption in OLS, we block bootstrap standard errors at the firm level in all models (Bertrand, Duflo, & Mullainathan, 2004) using 500 repetitions, which also accounts for correlations in error terms within firm that would otherwise be addressed through cluster corrections.

RESULTS

PFP Increases Average Stress Medication Usage

⁵ Neither the clogit or xtlogit commands in Stata will run due to “numerical overflow,” which represents a binomial coefficient exceeding the largest number representable in Stata.

Table 2 presents the results for our three measures of mental health problems: *benzo*, *ssri*, and *stress*. Column 1 shows that when a firm adopts P4P, existing employees become 0.29% more likely to use either benzodiazepine or SSRI medications. This represents an approximately 5.7% increase over base rate. Columns 2 and 3 suggest that this effect is split between benzodiazepines and SSRI medications, which not only treat depression but are also preferred for long-term anxiety treatment over benzodiazepines because of the latter's risk for dependency and abuse (Olfson, King, & Schoenbaum, 2015). The gain in *benzo* is 0.21% or 5.4% over base rate, while the *SSRI* is 0.16% or 11.3% over base rate. This implies that P4P generates 1,081 additional benzodiazepine and 824 SSRI prescription years in our data. Collectively, these results suggest that when firms adopt pay-for-performance, mental health issues increase among existing employees. Again, we caution that because P4P adoption is not random, we cannot make a strong causal argument. Conditional logit models (see Appendix), which suffer from the incidental parameters problem, produce similar albeit much larger effect sizes of 11%.

-----INSERT TABLE 2 HERE-----

As an alternative measure, we replace our P4P dummy variable with *pfppercent*, which is the logged percentage (0 to 100) of employees reported by the firm to use pay-for-performance. This variable applies differential treatment levels of P4P on firms based on the extent of the policy. The sample for these models is smaller because some firms that indicated using P4P did not report the magnitude of use. Results for these models are presented in columns 3 -- 6 of Table 2, and are consistent with the primary models.

PFP Motivates Attrition Based on Mental Health

Our individual fixed effects models estimate the average treatment effect on existing employees, but do not represent the average effect for the firm because they ignore differences in the mental health

of workers who leave or join following P4P. Even though remaining employees may suffer increased depression or anxiety, many with this health problem (or potential problems) may leave and be replaced by those without it. To address this, we implement difference-in-difference OLS models with firm fixed effects. In these models, the coefficient for p_{jt} represents the change in the probability that any worker at firm j in year t uses benzodiazapines or SSRIs following P4P adoption at the firm. As Lazear (2000) and Pierce et al. (2015) note, differences between coefficients in worker FE models and firm FE models indicate differences in unobservable traits. If P4P adoption does not motivate those with existing or potential mental health problems to leave at a disproportionately high rate, then the coefficients on p_{jt} should be identical to the worker fixed effect models. If, however, adoption motivates these workers to leave at a higher rate, then the coefficient should be smaller as new workers without mental health problems replace them.

Results for these models are presented in Table 3. Results are consistent with the worker fixed effects models, but the effect size of 0.23% is approximately 20% smaller than the coefficient in the worker fixed effects model, which yields several additional insights. We note, however, that these coefficients are not statistically distinguishable at traditional levels. However, these smaller estimates do suggest endogenous sorting by employees after P4P adoption. Since individual fixed effect models only estimate effects on those who stay, the larger coefficient in Table 2 indicates that those with either the propensity for or existing mental health problems leave the firm following P4P adoption and are replaced by those with lower propensity—consistent with predictions on endogenous sorting following P4P adoption. This suggests that the relationship between performance-based pay and mental health motivates job change in some workers following P4P adoption at a firm.⁶

⁶ Although P4P appears to influence job changes by certain employees, P4P is not correlated with higher average turnover, which suggests that some employees are less likely to leave after P4P adoption, possibly because it yields higher wages for them.

-----INSERT TABLE 3 HERE-----

Individual Differences in Mental Health Effects

We next examine several subsamples that might reveal mechanisms through which P4P could affect mental health. These subsamples, based on gender, age, and earnings, provide differential evidence for each mechanism, which we will discuss later. Again, we reiterate that our limited statistical power makes precise comparisons of group-based effect differences very difficult.

Gender differences: We repeat our worker fixed effect models separately for women and men in Table 4, and find no noticeable differences in mental health following P4P adoption between women and men who remain with the firm. Men and women who stay both appear to have higher medication usage after adoption, although parameter estimates are much less precise due to the smaller subsamples. Firm fixed effect models in Table 5, however, reveal a very different result across gender. Although the average female employee is no more likely to use medication after P4P, the average man has a large and precise increase of 0.45%. This difference between the gender gap in worker- and firm-fixed effect models suggests that women and men make different decisions about staying in firms that adopt P4P. The much smaller (zero) coefficient for women in firm fixed effect models is consistent with those women with existing or potential mental health problems leaving firms after they adopt P4P while being replaced by those without existing or potential anxiety or depression.

-----INSERT TABLES 4 AND 5 HERE-----

This result is consistent with extensive evidence that women are less likely to select performance-based pay or jobs with such pay when given an alternative (Gneezy & Rustichini, 2004; Niederle & Vesterlund, 2007; Dohmen & Falk, 2011; Barbulescu & Bidwell, 2013). What is puzzling is that the firm fixed effect coefficient for men is *larger* than in worker fixed effect models, which suggests that men with underlying mental health problems are more likely to stay with or join firms

after P4P adoption. Although this coefficient increase across models is not precise, it does imply that men with mental health problems are generally not leaving or avoiding P4P firms like their female counterparts. Why might this be the case? One reason might be that men are typically found to be more overconfident than women (Barber & Odean, 2001; Croson & Gneezy, 2009), such that some men might fail to understand or predict their ability to handle stress from performance-based-pay, or might view quitting a job because of stress as personally or social acceptable.

We also note that this differential sorting seems inconsistent with an alternative explanation—that outside labor market opportunities are primarily driving the decision to leave by those workers impacted by P4P adoption. Despite one of the most gender-progressive countries, women in Denmark still are less mobile than their male counterparts (Deding & Filges, 2010). We emphasize that the coefficient magnitudes across these models cannot be statistically distinguished through Wald tests, and must be interpreted cautiously.

Age differences: We also repeated our DiD models for workers in different age deciles: 20s, 30s, 40s, and over 50, and find that medication increases associated with P4P are primarily in older workers. Table 6, which presents separate regressions for each group, shows almost no change in mental health prescriptions among workers in their 20s and 40s. In contrast, workers in their 50s show an increase of 0.77%, an 8.9% increase over a base rate of 8.7% and an effect that is statistically distinguishable from the smaller effects in other workers ($p < .05$). Workers in their 30s show a smaller and less precise increase of 0.33%.

These results suggest several possible mechanisms behind the observed mental health effect. One possible reason is the stereotype that older workers might be less open to organizational change (Ng & Feldman, 2012). Restructuring and change might require different skill-sets, where the skills of older workers are disproportionately more likely to become obsolete (Bartel & Sicherman, 1993). Along these lines, studies have found that older workers are more likely to leave firms that adopt

organizational innovation and new technology (Aubert, Caroli, & Roger, 2006; Behaghel, Caroli, & Roger, 2014). Empirical work on more fundamental organizational change, such as plant downsizing and restructuring, report longer spells of sick leave among older workers after change (Vahtera, Kivimaki, Pentti, 1997).

The second possible reason is that older workers typically have less job mobility, which reduces the viability of leaving as a coping mechanism. The firm fixed effect regressions in Table 7 are consistent with this argument, showing an increased usage of 1.0% relative to individual fixed-effect models. In contrast, the effect for workers in the 30s, who have better job mobility, entirely disappears in firm fixed effect models. The effects in older workers are statistically distinguishable from those below age fifty ($p < .05$). Yet it is hard to reconcile these results with the gender-based ones based solely on job mobility. The more likely mechanism is that the same resistance to change that generates stress from P4P might also reduce willingness to change jobs to cope with the stress.

-----INSERT TABLES 6 AND 7 HERE-----

Stress increases tied to wage changes: One of the weaknesses of our data is that we cannot precisely observe who within the firm is receiving P4P. Although recent evidence suggests that P4P adoption might affect even those coworkers to which it is not applied (Lee & Puranam, 2016), such effects are likely to be smaller. To address this, we examine whether those workers most likely to suffer from P4P adoption—those whose wages decreased—have larger increases in mental health than their peers.

To do so, we divided those workers who were employed at P4P firms both before and after adoption in three ways and then identified two separate treatment effects for these subgroups in our worker fixed effects models. Model (1) in Table 9 shows that workers whose wages decreased following P4P adoption suffered all the medication increases, consistent with wage decreases serving as a mechanism through which P4P hurts mental health. Models (2) and (3) show similar results for

workers with below-mean and below-median wage changes. We note that wage change is endogenous, and might be influenced by mental health changes, thereby instead implying reverse causality. Similarly, the magnitudes of the coefficients for the two groups are not statistically distinguishable at conventional levels. So we are cautious in interpreting these results as demonstrating income as the defining mechanism. We also note that average income is actually *increasing* following P4P adoption for all classes of workers (see Appendix), which suggests that the many of the below-average workers with increased mental health in columns (2) and (3) are not suffering because of actual losses, but rather losses relative to their peers that might imply either social comparison costs or additional stress taken on to maintain income.

-----INSERT TABLE 8 HERE-----

Mental Health and Wage Loss Related to Post-P4P Turnover

The comparison between individual and firm fixed effect models suggests that workers leave in response to emerging mental health costs from P4P. In addition, the higher mental health problems emerging for lower-performing workers also suggests wage loss as a mechanism driving turnover. To more directly examine these paths, we directly test whether mental health problems and likely wage loss following P4P adoption are associated with the decision to leave the firm. We note considerable statistical problems with these tests. Turnover and severe stress (evidenced by medication) are endogenous competing hazards, and as our earlier results suggest, many workers likely leave to *avoid* medication usage. So we present these results purely as a robustness test. Similarly, observed wage changes might be correlated with other unobservable predictors of turnover following P4P adoption.

We define leaving the firm in any year if the employee enters unemployment, switches firms, or enters early retirement. We note that since we do not use the oldest workers, such retirement choices are early. For our mental health regressions, we regress this departure dummy on our three

medication dummies in separate regressions, also including our P4P indicator and their interaction. The results, presented in columns (1)-(3) of Table 9, show that stress and depression are strongly and precisely associated with turnover. Those taking medication are 5-9% more likely to leave in a given year. P4P usage appears slightly negatively associated with average turnover, although the parameter estimates are not precise ($p=.121$). This small effect may reflect countervailing effects of P4P on retention for employees of different abilities or performance levels (Shaw, 2015). Most importantly, the interaction term appears positive ($p=.067$). This indicates that mental health problems associated with P4P adoption are no less predictive of turnover than other mental health problems, and they may be slightly larger. These results are consistent with both turnover as a coping mechanism for stress and depression both generally and as a response to P4P adoption.

We also examine whether the introduction of P4P is associated with higher turnover for lower performing workers. Although ideally we would correlate actual wage changes following P4P adoption with the probability of turnover, this is problematic because of our annual data. Those who leave in the first year would likely report reduced total wages because of a shorter time at the firm. Similar to our earlier analysis on performance and mental health, we take only those employees in firms that adopted P4P and worked for two years prior to P4P adoption. We designate those with an below average raise (change in wages) in the year prior to P4P adoption as low performers who would likely suffer wage loss under P4P. For this much smaller sample, we present regressions in columns (4) and (5) that regress departure on P4P and its interaction with below-average wage trajectory (the main effect is absorbed by fixed effects). Both columns show that for this sample of longer-term employees, P4P actually sharply increases turnover by 6-12%. But the interaction effect shows an additional 2-4% increase for low performers. This result is consistent with a lower pay mechanism as a path through which P4P might affect turnover and sorting.

-----INSERT TABLE 9 HERE-----

Extensions

Other health implications: Although SSRIs and benzodiazepines are the most direct measure of treated stress and depression, mental health problems may also manifest themselves through other physical ailments that may require medication. We examine three such medication classes: diabetes (ATC-10), beta-blockers (for high blood pressure), and statins for high cholesterol.⁷ Although there is little biological evidence that stress or depression might increase long-term blood pressure or cholesterol, they might motivate activities such as diet changes or inactivity that could indirectly reduce these health measures. In contrast, stress has been shown to directly increase blood glucose through the release of epinephrine, glucagon, growth hormone, and cortisol.⁸ We repeat our worker fixed effect models, and find no evidence of increased diabetes, blood pressure, or statin usage (see Appendix).

Time trends in effects: To examine how quickly increases in mental health medication usage occur following P4P implementation, we implement a “lags and leads” model (Angrist & Pischke, 2008), where we estimate a separate treatment coefficient for each year before and after P4P adoption, considering the year before adoption as the baseline. These results, presented in Figure 3 with \pm two standard errors, indicate that although a small treatment effect occurs initially, medication usage grows with each subsequent year. This is consistent with either individuals either delayed effects for some workers or else affected individuals delaying treatment. It is impossible for us, however, to separate these possible explanations using our data. This increasing treatment effect also suggests that individuals who stay at the firm fail to psychologically adapt to P4P, at least in

⁷ Although ADHD medications, which are often used as cognitive enhancers in the workplace (Greely, Campbell, Shakian, Harris, & Kessler, 2018), might also increase following P4P adoption, their use is extremely rare in our data (0.01%).

⁸ <https://dtc.ucsf.edu/types-of-diabetes/type2/understanding-type-2-diabetes/how-the-body-processes-sugar/blood-sugar-stress/>

ways that would allow them to stop usage of benzodiazepines or SSRIs. Our lags and leads models also show no evidence that our data violate the crucial parallel trends assumption in difference-in-differences models.

-----INSERT FIGURE 3 HERE-----

Other HR policy changes: As we noted earlier, firms often adopt human resource practices as systems of policies (Arthur, 1994; Huselid, 1995; Ichniowski & Shaw, 1999; Laursen & Foss, 2003, 2014). Since the DISKO2 survey asks managers about the adoption date of multiple HR practices, we repeated our primary difference-in-difference model with worker fixed effects on the combined *stress* variable. We test the nine other policies listed in Questions 8-9 (see Figure 1): interdisciplinary workgroups, formal delegation of quality control, systems for collecting proposals from employees, planned job rotation, delegation of responsibility, autonomous groups, integration of functions (e.g., sales, production), telework/distance work, and wages according to qualifications or function (scaled wages).

The results, presented in Figure 4, show no other statistically significant results. The largest and most precisely estimated coefficients are those for job rotation and delegation, but neither is significant at the 10% level. An interesting result is that the coefficient for “systems for collecting proposals from employees” is positive, which is inconsistent with the large literature espousing the importance of employee voice (Burris, 2002; Detert & Edmondson, 2011; Detert & Burris, 2007; Morrison, 2011), but this parameter estimate is also imprecise. In addition, we tested for complementarities between these policies and P4P adoption through a triple differences model that interacts P4P adoption with one of the nine other policies (see Appendix). Of these nine, only telework has negative and marginally significant effect ($p=.09$), but given the multiple comparisons tested in this figure, it is hard to be confident that this is a true effect and not just a false positive from nine separate regressions.

-----INSERT FIGURE 4 HERE-----

Economic conditions: We additionally test whether the increases in mental health problems associated with P4P were more severe under worse economic conditions, which might suppress pay and limit job mobility. To do this, we implement triple-difference models where the interaction of PFP_{jt} and the economic condition in year t represents the increased effect of P4P based on two measures: unemployment rate and GDP growth rate. These models found no marginal effect for either unemployment rate ($b=-.0003$, $p=.74$) or GDP growth ($-.0007$, $p=.567$).

Counterfactual Simulations: To ensure that our results are not mechanically generated, we conduct a standard difference-in-difference placebo test (Bertrand et al. 2004; Pierce et al., 2015; Staats, Dai, Hofmann, & Milkman, 2016) by randomly assigning each firm another firm's P4P adoption date (or no date) and rerunning our individual fixed effect models. We ran 1,000 placebo regressions and plot coefficient estimates in Figure 5 with 95% confidence intervals. Our real data's estimated treatment effect is represented in black, and is far more precise and larger than most of the placebo trials.

-----INSERT FIGURE 5 HERE-----

DISCUSSION

Consistent with prior theory from multiple fields, our results suggest that the mental health of employees may indeed suffer when firms adopt pay-for-performance systems. Our primary difference-in-difference models find that mental health medication usage for existing employees increases by 5.6% over base rate after a firm adopts P4P, relative to other firms in the same year. Although the immediate relationship between pay-for-performance stress is known, our results additionally show that this short-term stress can develop into serious diagnosable mental health problems. Furthermore, we provide this first evidence using over a thousand firms and hundreds of

thousands of employees, with objective medical data that does not suffer from self-response bias. If one were to extrapolate our estimates from Denmark to a much larger country such as the United States, they would imply up to an additional half million prescription years from pay-for-performance adoption. Consequently, this paper represents an important extension of anecdotal and smaller scale studies—there are real long-term psychological costs of pay-for-performance with significant financial and social welfare implications.

In addition, the smaller parameter estimates for firm fixed effects models suggest that the mental health effects of pay-for-performance also motivate workers to change job following P4P adoption. Our results are consistent with workers with unobserved mental health vulnerability being replaced by those with lower vulnerability following P4P adoption, which suggests workers strategically adapt to compensation policy changes through costly job change. This is consistent with arguments by Gerhart and Fang (2014) that employee sorting should be a much larger consideration in firm considerations of pay-for-performance adoption. Unfortunately, because the DISKO2 survey does not cover all firms, we cannot identify whether the firms to which these departing workers move also use pay-for-performance.

Implications for Theory

What do our results imply for future theoretical advances on the implications of performance-based pay for employee outcomes? Our results collectively imply a framework similar to Figure 6, which presents both the observed links between P4P and mental health as well as the remaining questions that our study cannot answer. Pay-for-performance appears to have a direct positive relationship with anxiety and depression in our data, but we are unable to extract all the mechanisms that might drive that relationship. Our results also imply, as in Figure 6, that both the actual and potential severe mental health problems from P4P will lead some workers to switch jobs.

As the figure notes, some workers will leave before they need or choose medication, as we observe in our results. Others may leave only after their problems become so severe as to require medication.

-----INSERT FIGURE 6 HERE-----

Our inability to observe pre-medication depression and anxiety prevents us from unraveling the mechanisms that might explain why some would leave and others would stay. One economic mechanism, which is the availability of outside labor market options, is supported by our results on older workers, who have decreased outside job options compared to younger workers. Without good outside job options, they may be “trapped” in a job with increased risk and pay uncertainty, which may generate stress or depression. But older people’s reduced adaptability to changing work or life environments suggests that job mobility is a complex combination of economic, social, and psychological forces. A better understanding of these mechanisms could help clarify the unique challenges facing older workers under P4P schemes, particularly as many countries extend the minimum age for fully pension or social security.

Our results on gender suggest that labor markets may be far less important in determining turnover than more social and psychological processes. Although women and men who remain with firms after P4P adoption suffer similar increases in mental health problems, they clearly cope with them differently. Women appear to selectively leave or join firms following P4P adoption based on their mental health effects from performance-based pay, while men fail to do so. This seems to run counter to an explanation of outside job options, since women (like older workers) typically have *less* labor mobility than male counterparts, even in gender-progressive Denmark (Deding & Filges, 2010). The unobservable difference in effects across unemployment rates also supports this limited role. These results collectively suggest the importance of psychological mechanisms with known differences across gender, such as overconfidence or social expectations, that might explain these differences. Given that the large-scale anonymous data in our study cannot identify this decision

process, qualitative or survey-based research may be needed to uncover this process to help guide theory on individual differences in adjustment to organizational change.

Additionally, our results imply an indirect path from P4P to mental health problems through wage loss. Low performers in our sample are more likely to both suffer mental health problems and leave the firm following P4P adoption, but we cannot separate the relative importance of this indirect path from the more direct path through other unobservable mechanisms. Similarly, we cannot compare their relative importance in driving turnover, which might also result from a rational response to the reduced wages under P4P, independent of mental health issues.

The framework in Figure 6 raises as many questions as answers, and reflects the relative strengths and weaknesses of our data. Our data and analysis raise confidence over prior work that P4P drives *severe and economically important* mental health problems unobservable in prior work, even as it cannot observe many less severe cases. Similarly, our results hint at likely classes of mechanisms behind this relationship across a large array of firms, but cannot dig deeply into the relative importance of specific mechanisms in driving observable outcomes. Given these limitations, what type of empirical study would be needed to better define such a model?

The data necessary to advance such a model would require many of the elements in our own data, but with much richer medical, pay, and self-response components that might review mechanisms. First, researchers would need sufficient cross-sectional and longitudinal variation in the adoption of P4P across individuals, even within job function. This would suggest data that either span many separate firms or else one or more large firms with location-specific and time varying P4P policies. Ideally, such P4P policies would enjoy some quasi-random shock to address the endogenous pay policy adoption challenges in our data. Second, these data would need to be linked to health claims data that would provide a more comprehensive mental health picture (i.e., counseling) than our medication data. In the United States, firms are barred from possessing such

data about employees, so the researcher would need to separately acquire employer and health plan data and then link and de-identify them under HIPAA compliant procedures (see Gubler et al., 2018 for an example). Ultimately, rich evidence on mechanisms would seem to require complementary survey data on the experiences, emotions, and preferences of the employees in the sample.

Where might one acquire such a comprehensive data panel? One starting point would be a major employment-based health insurance or benefits provider. If such a firm believed the study could help improve health risk models, then it might solicit client firms to participate in combining and anonymizing data through the third-party researcher. Client firms might possess their own job satisfaction and climate surveys that could be used to tease out possible mechanisms. The difficult and legally-fraught process for building such data reflects why our study, even with its many weaknesses, represents an important empirical contribution on the effects of pay-for-performance.

Implications for Managers

What are the implications of our findings for firms? If mental health in some workers indeed degrades under P4P, then this represents an important cost to firm productivity that must be considered in compensation policy design. Poor health, whether mental or physical, has been widely linked to lower worker productivity (Currie & Madrian, 1999; Thayer, Newman, & McClain, 1994; Christian, Eisenkraft, & Kapadia, 2015; Gubler, Larkin, & Pierce, 2018). It is impossible for us to weigh possible motivational gains versus mental health costs in our data, so we are hopeful future data can better measure net benefits to performance. An ideal study would link individual productivity and mental health data to compensation changes at the firm, but such data combinations are rare. Firms in both Denmark and the United States are typically legally barred from holding employee medical data, and medical data such as ours typically cannot be linked to individual data by researchers. Rare examples such as Gubler, Larkin, and Pierce (2018), which relies

on policy changes at one small firm, typically do not have the statistical power to identify the net performance implications of policies that affect health.

Key Limitations

Our paper has several key limitations beyond our inability to tease out specific mechanisms. First, we are careful to note that our paper cannot provide causal evidence that P4P causes mental health problems. P4P adoption is clearly endogenous (see Gooderham, Fenton-O’Creevy, Croucher, & Brooks, 2018), and while we can be confident that employee mental health is not driving P4P adoption (i.e., reverse causality), there are undoubtedly endogenous selection biases in our analysis based on omitted variables. Despite these causal inference limitations, we believe our unique combination of precise wage and medication data for a large sample of firms and employees richly complements the existing literature, and sparks additional work that might link employee medical and health data with productivity, operations, and human resource policy data from firms (Dahl, 2011; Gubler et al. 2018; Gubler & Pierce, 2014; ten Brummelhuis, Rothbard, & Uhrich, 2017).

Second, as we’ve noted earlier in the paper, the low base rate of SSRI and benzodiazepine usage in our sample means that our 4.5% observed effect is only responsible for about 2,000 years of prescriptions in our data. Some readers will view this coefficient as inconsequential in a population of 318,717 workers. We disagree with such assessment. In and of itself, a marginal increase of 1,081 benzodiazepine and 824 SSRI years of prescriptions represents significant economic and public health costs to society. Furthermore, this is equivalent to over half a million additional prescriptions in a country the size of the United States. Depression and anxiety disorders are serious diseases with high mortality risk (Bostwick & Pankratz, 2000), such that small effects matter. In this way, our paper complements existing work that shows how P4P adoption can generate significant physical health problems (DeVaro & Heywood, 2017; Foster & Rosenzweig, 1994; Freeman & Kleiner, 2005; Böckerman, Bryson, & Ilmakunnas, 2012; Artz & Heywood, 2015).

Even more important, if these severe cases represent the tip of a much larger effect, unobservable in our data, the costs could be far more severe. Performance-based pay could be associated with many more depression and anxiety cases either treated through psychotherapy or remaining untreated.

Additionally, we reiterate that despite the size of our panel dataset, our study does not have high enough statistical power to estimate extremely precise parameters. Although most of our estimates meet traditional significance levels for full samples, our subsample estimates are far less precise. This is a challenge with doing panel research with rare (but important) events, particularly where observations are clustered in groups. So we caution readers to view our paper as initial but not definitive evidence.

Finally, it is important to note that our study cannot comment on the overall net benefit from implementing pay-for-performance in firms. Although we have demonstrated an increase in mental health problems—a real cost to individuals, firms, and society—we are not comparing this against possible gains from P4P. Certainly, in our data we observe wages increasing in P4P firms, so it is hard to weigh these and other benefits against the small portion of the population with medically-documented anxiety and depression. So we caution the reader to view the evidence here as documenting one important factor in a complex calculation of the net impact of P4P policies. But we note that the extensive costs of depression documented by Greenberg et al. (2015) are daunting. Depression and other mental health problems generate costs far beyond absenteeism and workplace productivity, also permeating beyond the individual to impact their family and broader social network. Given our estimated magnitude and likely severity of these effects, the costs we observe are unlikely to be negligible.

References

- Allan, J., Bender, K., & Theodossiou, I. 2017. Performance pay and stress: an experimental study. Unpublished working paper. University of Aberdeen: <https://core.ac.uk/download/pdf/82969755.pdf>

- Angrist, J., & Pischke, J. 2008. *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton University Press.
- Arthur, J. B. 1994. Effects of human resource systems on manufacturing performance and turnover. *Academy of Management Journal*, 37(3): 670-687.
- Artz, B., & Heywood, J. S. 2015. Performance pay and workplace injury: panel evidence. *Economica*, 82(s1): 1241-1260.
- Aubert, P., Caroli, E., & Roger, M. 2006. New technologies, organization and age: Firm-level evidence. *The Economic Journal*, 1: F73-F93.
- Baker, G. 2000. The use of performance measures in incentive contracting. *The American Economic Review*, 90(2) : 415-420.
- Bandiera, O., Barankay, I., & Rasul, I. 2005. Social preferences and the response to incentives: Evidence from personnel data. *The Quarterly Journal of Economics*, 120(3): 917-962.
- Barber, B., & Odean, T. 2001. Boys will be boys: Gender, overconfidence, and common stock investment. *The Quarterly Journal of Economics*, 116(1): 261-292.
- Barbulescu, R., & Bidwell, M. 2013. Do women choose different jobs from men? Mechanisms of application segregation in the market for managerial workers. *Organization Science*, 24(3): 737-756.
- Bartel, A. P., & Sicherman, N. 1993. Technological change and retirement decisions of older workers. *Journal of Labor Economics*, 1:162-183.
- Begley, T., & Czajka, J. 1993. Panel analysis of the moderating effects of commitment on job satisfaction, intent to quit, and health following organizational change. *Journal of Applied Psychology*, 78(4): 552-556.
- Behaghel, L., Caroli, E., & Roger, M. 2014. Age-biased technical and organizational change, training and employment prospects of older workers. *Economica*, 81: 368–389.
- Benabou, R., & Tirole, J. 2003. Intrinsic and extrinsic motivation. *The Review of Economic Studies*, 70(3): 489-520.
- Bender, K., & Theodossiou, I. 2013. The unintended consequences of the rat race: the detrimental effects of performance pay on health. *Oxford Economic Papers*, 66(3), 824-847.
- Bertrand, M., Duflo, E., & Mullainathan, S. 2004. How much should we trust differences-in-differences estimates?. *The Quarterly Journal of Economics*, 119(1), 249-275
- Blumberg, S. J., Clarke, T. C., & Blackwell, D. L. 2015. *Racial and ethnic disparities in men's use of mental health treatments*. US Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Health Statistics.
- Böckerman, P., Bryson, A., & Ilmakunnas, P. 2012. Does high involvement management improve worker wellbeing?. *Journal of Economic Behavior & Organization*, 84(2): 660-680.
- Bolino, M. C., & Grant, A. M. 2016. The bright side of being prosocial at work, and the dark side, too: a review and agenda for research on other-oriented motives, behavior, and impact in organizations. *Academy of Management Annals*, 10(1): 599-670.
- Booth, A. L., & Frank, J. 1999. Earnings, productivity, and performance-related pay. *Journal of Labor Economics*, 17(3): 447-463.
- Bostwick, J., & Pankratz, V. 2000. Affective disorders and suicide risk; a reexamination. *American Journal of Psychiatry*, 157(12): 1925-1932. Bridges, S., & Disney, R. 2010. Debt and depression. *Journal of Health Economics*, 29(3): 388-403.
- Burks, S. V., Carpenter, J. P., Goette, L., & Rustichini, A. 2009. Cognitive skills affect economic preferences, strategic behavior, and job attachment. *Proceedings of the National Academy of Sciences*, 106(19): 7745-7750.
- Burris, E. R. 2012. The risks and rewards of speaking up: Managerial responses to employee voice. *Academy of Management Journal*, 55(4), 851-875.

- Cadsby, C. B., Song, F., & Tapon, F. 2007. Sorting and incentive effects of pay-for-performance: An experimental investigation. *Academy of Management Journal*, 50(2): 387-405.
- Cadsby, C. B., Song, F., & Tapon, F. 2016. The impact of risk-aversion and stress on the incentive effect of performance-pay. *Experiments in Organizational Economics*: 189-227. Emerald Group Publishing Limited.
- Chan, T. Y., Li, J., & Pierce, L. 2014a. Compensation and peer effects in competing sales teams. *Management Science*, 60(8): 1965-1984.
- Chan, T. Y., Li, J., & Pierce, L. 2014b. Learning from peers: Knowledge transfer and sales force productivity growth. *Marketing Science*, 33(4): 463-484.
- Charness, G., Masclet, D., & Villeval, M. C. 2013. The dark side of competition for status. *Management Science*, 60(1): 38-55.
- Christian, M.S., Eisenkraft, N., & Kapadia, C. 2015. Dynamic associations among somatic complaints, human energy, and discretionary behaviors experiences with pain fluctuations at work. *Administrative Science Quarterly*, 60(1): 66-102.
- Cornelissen, T., Heywood, J. S., & Jirjahn, U. 2011. Performance pay, risk attitudes and job satisfaction. *Labour Economics*, 18(2): 229-239.
- Croson, R. & Gneezy, U., 2009. Gender differences in preferences. *Journal of Economic Literature*, 47(2): 448-74.
- Currie, J. & Madrian, B.C. 1999. Health, health insurance and the labor market: Chapter 50. In O. C. Ashenfelter and D. Card (Eds.), *Handbook of Labor Economics*. 3309–3416. Elsevier.
- Dahl, M. S. 2011. Organizational change and employee stress. *Management Science*, 57(2): 240-256.
- Dahl, M. S., Nielsen, J., & Mojtabai, R. 2010. The effects of becoming an entrepreneur on the use of psychotropics among entrepreneurs and their spouses. *Scandinavian Journal of Public Health*, 38(8): 857-863.
- Dahl, M. S., & Sorenson, O. 2010. The social attachment to place. *Social Forces*, 89(2): 633-658.
- Davis, M. 2016. Pay matters: the piece rate and health in the developing world. *Annals of Global Health*, 82: 858-871.
- Dawling, P., & Falk, A. 2011. Performance pay and multidimensional sorting: Productivity, preferences, and gender. *The American Economic Review*, 101(2), 556-590.
- Deding, M., & Filges, T. 2010. Geographical mobility of Danish dual-earner couples—The relationship between change of job and change of residence. *Journal of Regional Science*, 50(2), 615-634.
- Detert, J. R., & Burris, E. R. 2007. Leadership behavior and employee voice: Is the door really open?. *Academy of Management Journal*, 50(4), 869-884.
- Detert, J. R., & Edmondson, A. C. 2011. Implicit voice theories: Taken-for-granted rules of self-censorship at work. *Academy of Management Journal*, 54(3), 461-488.
- Dohmen, T., & Falk, A. 2011. Performance pay and multidimensional sorting: Productivity, preferences, and gender. *The American Economic Review*, 101(2): 556-590.
- Devaro, J., & Heywood, J. S. 2017. Performance pay and work-related health problems: A longitudinal study of establishments. *ILR Review*, 70(3): 670-703.
- Drago, R., & Garvey, G. T. 1998. Incentives for helping on the job: Theory and evidence. *Journal of Labor Economics*, 16(1): 1-25.
- Edelman, B., & Larkin, I. 2014. Social comparisons and deception across workplace hierarchies: Field and experimental evidence. *Organization Science*, 26(1): 78-98.
- Eriksson, T., & Lausten, M. 2000. Managerial pay and firm performance — Danish evidence. *Scandinavian Journal of Management*, 16(3): 269-286.
- Eriksson, T., & Villeval, M. C. 2008. Performance-pay, sorting and social motivation. *Journal of Economic Behavior and Organization*, 68(2): 412-421.

- Fehr, E., & Schmidt, K. 1999. A theory of fairness, competition, and cooperation. *The Quarterly Journal of Economics*, 114: 817-868.
- Feldman, E., Gartenberg, C., & Wulf, J. 2018. Pay inequality and corporate divestitures. *Strategic Management Journal*. Forthcoming.
- Festinger, L. 1954. A theory of social comparison processes. *Human Relations*, 7(2): 117-140.
- Frank, D. H., & Obloj, T. 2014. Firm-specific human capital, organizational incentives, and agency costs: Evidence from retail banking. *Strategic Management Journal*, 35(9): 1279-1301.
- Frederiksen, A., & Westergaard-Nielsen, N. 2007. Where did they go? Modelling transitions out of jobs. *Labour Economics*, 14(5): 811-828.
- Freeman, R. B., & Kleiner, M. M. 2005. The last American shoe manufacturers: Decreasing productivity and increasing profits in the shift from piece rates to continuous flow production. *Industrial Relations: A Journal of Economy and Society*, 44(2): 307-330.
- Frey, B. 1997. *Not Just for the Money*. Edward Elgar Publishing.
- Frey, B. S., & Jegen, R. 2001. Motivation crowding theory. *Journal of Economic Surveys*, 15(5): 589-611.
- Foss, N., & Laursen, K. 2005. Performance pay, delegation and multitasking under uncertainty and innovativeness: An empirical investigation. *Journal of Economic Behavior & Organization*, 58(2), 246-276.
- Foster, A. D., & Rosenzweig, M. R. 1994. A test for moral hazard in the labor market: Contractual arrangements, effort, and health. *The Review of Economics and Statistics*. 213-227.
- Garcia, S. M., & Tor, A. 2007. Rankings, standards, and competition: Task vs. scale comparisons. *Organizational Behavior and Human Decision Processes*, 102(1): 95-108.
- Garcia, S. M., Tor, A., & Gonzalez, R. 2006. Ranks and rivals: A theory of competition. *Personality and Social Psychology Bulletin*, 32(7): 970-982.
- Gartenberg, C., & Wulf, J. 2017. Pay harmony? Social comparison and performance compensation in multibusiness firms. *Organization Science*, 28(1): 39-55.
- Gartenberg, C., & Wulf, J. 2018. Islands of equality: Competition and pay inequality within and across firm boundaries. Unpublished Working Paper. Wharton Business School.
- Gerhart, B., & Fang, M. 2014. Pay for (individual) performance: Issues, claims, evidence and the role of sorting effects. *Human Resource Management Review*, 24(1): 41-52.
- Gerhart, B., & Rynes, S. 2003. *Compensation: Theory, evidence, and strategic implications*. SAGE Publications.
- Gomez-Mejia, L. R., & Welbourne, T. M. 1988. Compensation strategy: An overview and future steps. *People and Strategy*, 11(3): 173.
- Gneezy, U., Meier, S., & Rey-Biel, P. 2011. When and why incentives (don't) work to modify behavior. *The Journal of Economic Perspectives*, 25(4): 191-209.
- Gneezy, U., & Rustichini, A. 2004. Gender and competition at a young age. *The American Economic Review*, 94(2): 377-381.
- Greely, H., Sahakian, B., Harris, J., Kessler, R. C., Gazzaniga, M., Campbell, P., & Farah, M. J. 2008. Towards responsible use of cognitive-enhancing drugs by the healthy. *Nature*, 456(7223): 702.
- Greenberg, P., Fournier, A., Sisisky, t., Pike, C., & Kessler, R. 2015. The economic burden of adults with major depressive disorder in the United States (2005 and 2010). *Journal of Clinical Psychiatry*, 76(2): 155-162.
- Gubler, T., Larkin, I., & Pierce, L. Forthcoming 2018. Doing well by making well: The impact of corporate wellness programs on employee productivity. *Management Science*.
- Gubler, T., & Pierce, L. 2014. Healthy, wealthy, and wise: Retirement planning predicts employee health improvements. *Psychological Science*, 25(9): 1822-1830.
- Gürtler, O. 2008. On sabotage in collective tournaments. *Journal of Mathematical Economics*, 44(3): 383-393.

- Hamilton, B. H., Nickerson, J. A., & Owan, H. 2003. Team incentives and worker heterogeneity: An empirical analysis of the impact of teams on productivity and participation. *Journal of Political Economy*, 111(3): 465-497.
- Hitt, M. A., Bierman, L., Shimizu, K., & Kochhar, R. 2001. Direct and moderating effects of human capital on strategy and performance in professional service firms: A resource-based perspective. *Academy of Management Journal*, 44(1): 13-28.
- Hölmstrom, B. 1979. Moral hazard and observability. *The Bell Journal of Economics*: 74-91.
- Holmstrom, B., & Milgrom, P. 1991. Multitask principal-agent analyses: Incentive contracts, asset ownership, and job design. *Journal of Law, Economics, and Organization*, 7: 24-52.
- Huselid, M. A. 1995. The impact of human resource management practices on turnover, productivity, and corporate financial performance. *Academy of Management Journal*, 38(3): 635-672.
- Ichniowski, C., & Shaw, K. 1999. The effects of human resource management systems on economic performance: An international comparison of US and Japanese plants. *Management Science*, 45(5): 704-721.
- Jensen, M. C., & Murphy, K. J. 1990. Performance pay and top-management incentives. *Journal of Political Economy*, 98(2): 225-264.
- Kahneman, D., Knetsch, J. L., & Thaler, R. H. 1991. Anomalies: The endowment effect, loss aversion, and status quo bias. *The Journal of Economic Perspectives*, 5(1): 193-206.
- Katz, E. 2001. Bias in conditional and unconditional fixed effects logit estimation. *Political Analysis*, 9(4): 379-384.
- Kim, T. Y., Weber, T. J., Leung, K., & Muramoto, Y. 2009. Perceived fairness of pay: The importance of task versus maintenance inputs in Japan, South Korea, and Hong Kong. *Management and Organization Review*, 6(1): 31-54.
- Knight, A., Menges, J., & Bruch, H. 2017. Organizational affective tone: A meso perspective on the origins and effects of consistent affect in organizations. *Academy of Management Journal*.
- Lancaster, T. 2000. The incidental parameter problem since 1948. *Journal of Econometrics*, 95(2): 391-413.
- Larkin, I. 2014. The cost of high-powered incentives: Employee gaming in enterprise software sales. *Journal of Labor Economics*, 32(2): 199-227.
- Larkin, I., & Leider, S. 2012. Incentive schemes, sorting, and behavioral biases of employees: Experimental evidence. *American Economic Journal: Microeconomics*, 4(2): 184-214.
- Larkin, I., & Pierce, L. 2015. Compensation and employee misconduct: the inseparability of productive and counterproductive behavior in firms. In D. Palmer, R. Greenwood & K. Smith-Crowe (Eds.), *Organizational wrongdoing: Key perspectives and new directions*. 1-27. Cambridge, UK: Cambridge University Press.
- Larkin, I., Pierce, L., & Gino, F. 2012. The psychological costs of pay-for-performance: Implications for the strategic compensation of employees. *Strategic Management Journal*, 33(10): 1194-1214.
- Laursen, K., & Foss, N. J. 2003. New human resource management practices, complementarities and the impact on innovation performance. *Cambridge Journal of Economics*, 27(2): 243-263.
- Laursen, K., & Foss, N. J. 2014. Human resource management practices and innovation. In *Oxford Handbook of Innovation Management* (pp. 506-529). Oxford University Press.
- Lazear, E. P. 1986. Salaries and piece rates. *Journal of Business*. 405-431.
- Lazear, E. P. 1999. Culture and language. *Journal of Political Economy*, 107(S6): S95-S126.
- Lazear, E. P. 2000. Performance pay and productivity. *American Economic Review*, 90(5): 1346-1361.
- Lee, E., & Puranam, P. 2016. The implementation imperative: Why one should implement even imperfect strategies perfectly. *Strategic Management Journal*, 37(8): 1529-1546.
- Morrison, E. W. 2011. Employee voice behavior: Integration and directions for future research. *Academy of Management Annals*, 5(1), 373-412.

- Nickerson, J. A., & Zenger, T. R. 2008. Envy, comparison costs, and the economic theory of the firm. *Strategic Management Journal*, 29(13): 1429-1449.
- Niederle, M., & Vesterlund, L. 2007. Do women shy away from competition? Do men compete too much?. *Quarterly Journal of Economics*, 122(3): 1067-1101.
- Ng, T. W., & Feldman, D. C. 2012. Evaluating six common stereotypes about older workers with meta-analytical data. *Personnel Psychology*, 65(4): 821-858.
- Nyberg, A. J., Pieper, J. R., & Trevor, C. O. 2016. Pay-for-performance's effect on future employee performance: Integrating psychological and economic principles toward a contingency perspective. *Journal of Management*, 42(7): 1753-1783.
- Obloj, T., & Zenger, T. 2017. Organization design, proximity, and productivity responses to upward social comparison. *Organization Science*, 28(1): 1-18.
- Olson M., King M., & Schoenbaum M. 2015. Benzodiazepine use in the United States. *JAMA Psychiatry*, 72(2):136-142.
- Pfeffer, J., & Carney, D. Forthcoming 2018. The economic evaluation of time can cause stress. *Academy of Management Discoveries*.
- Pierce, L., Dahl, M.S. & Nielsen, J. 2013. In sickness and in wealth: Psychological and sexual costs of income comparison in marriage. *Personality and Social Psychology Bulletin*, 39(3): 359-374.
- Pierce, L., Snow, D. C., & McAfee, A. 2015. Cleaning house: The impact of information technology monitoring on employee theft and productivity. *Management Science*, 61(10): 2299-2319.
- Pierce, L., Wang, L., & Zhang, D. 2019. Peer bargaining and productivity in teams: Gender and the inequitable division of pay. Unpublished working paper.
- Prendergast, C. 1999. The provision of incentives in firms. *Journal of Economic Literature*, 37(1): 7-63.
- Ryan, R. M., & Deci, E. L. 2000. Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist*, 55(1): 68.
- Rynes, S. L., Gerhart, B., & Parks, L. 2005. Personnel psychology: Performance evaluation and pay-for-performance. *Annual Review of Psychology*, 56: 571-600.
- Shaw, J. D. 2015. Pay dispersion, sorting, and organizational performance. *Academy of Management Discoveries* 1(2): 165-179.
- Staats, B. R., Dai, H., Hofmann, D., & Milkman, K. L. 2016. Motivating process compliance through individual electronic monitoring: An empirical examination of hand hygiene in healthcare. *Management Science*, 63(5), 1563-1585.
- Sweet, E., Nandi, A., Adam, E. K., & McDade, T. W. 2013. The high price of debt: Household financial debt and its impact on mental and physical health. *Social Science and Medicine*, 91: 94-100.
- ten Brummelhuis, L., Rothbard, N., & Uhrich, B. Forthcoming 2017. Beyond nine to five: Is working to excess bad for health?. *Academy of Management Discoveries*, 3(3): 262-283.
- Thayer R.E., Newman J.R., & McClain T.M. 1994. Self-regulation of mood: strategies for changing a bad mood, raising energy, and reducing tension. *Journal of Personality and Social Psychology*, 67(5): 910.
- Trevor, C. O., Reilly, G., & Gerhart, B. 2012. Reconsidering pay dispersion's effect on the performance of interdependent work: Reconciling sorting and pay inequality. *Academy of Management Journal*, 55(3): 585-610.
- Vahtera J., Kivimäki, M., & Pentti, J., 1997. Effect of organisational downsizing on health of employees. *The Lancet*, 350(9085): 1124-1128.
- Vroom, V.H., 1964. *Work and motivation*. New York: Wiley & Sons.
- World Health Organization. 2014. *Social determinants of mental health*. World Health Organization.
- Zenger, T. R. 1994. Explaining organizational diseconomies of scale in RandD: Agency problems and the allocation of engineering talent, ideas, and effort by firm size. *Management Science*, 40(6): 708-729.

Figure 1: DISKO2 Pay-for-Performance Questions

8. Does the firm make use of some of the following ways of planning the work and paying the employees
(Mark with an X at each example)

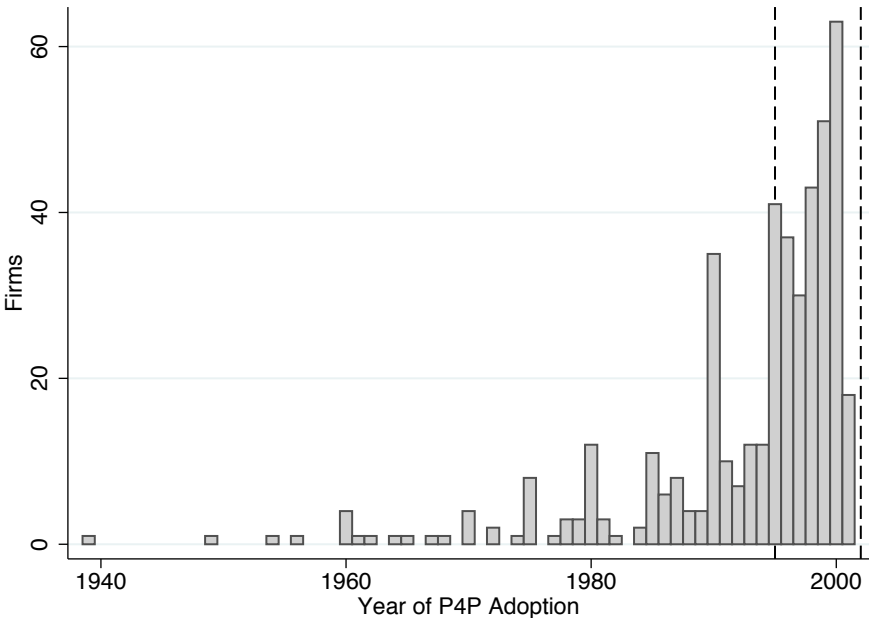
	Yes	No	Don't know
Interdisciplinary workgroups			
Quality circles/groups (Formal delegation of quality control)			
Systems for collecting proposals from employees			
Planned job rotation			
Delegation of responsibility			
Autonomous groups			
Integration of functions (e.g. sales, production)			
Tele work/distance work (formal agreements)			
Wages according to qualifications or function			
Performance related pay (not piece-rate pay)			

9. When were these measures introduced and how many of the employees are included (app. no) ?
(is only to be answered if one or more questions were marked with x in question 8)

	When? (year)	App. Share of employees included
Interdisciplinary workgroups	_____	_____ %
Quality circles/groups (Formal delegation of quality control)	_____	_____ %
Systems for collecting proposals from employees	_____	_____ %
Planned job rotation	_____	_____ %
Delegation of responsibility	_____	_____ %
Autonomous groups	_____	_____ %
Integration of functions (e.g. sales, production)	_____	_____ %
Tele work/distance work (formal agreements)	_____	_____ %
Wages according to qualifications or function	_____	_____ %
Performance related pay (not piece-rate pay)	_____	_____ %

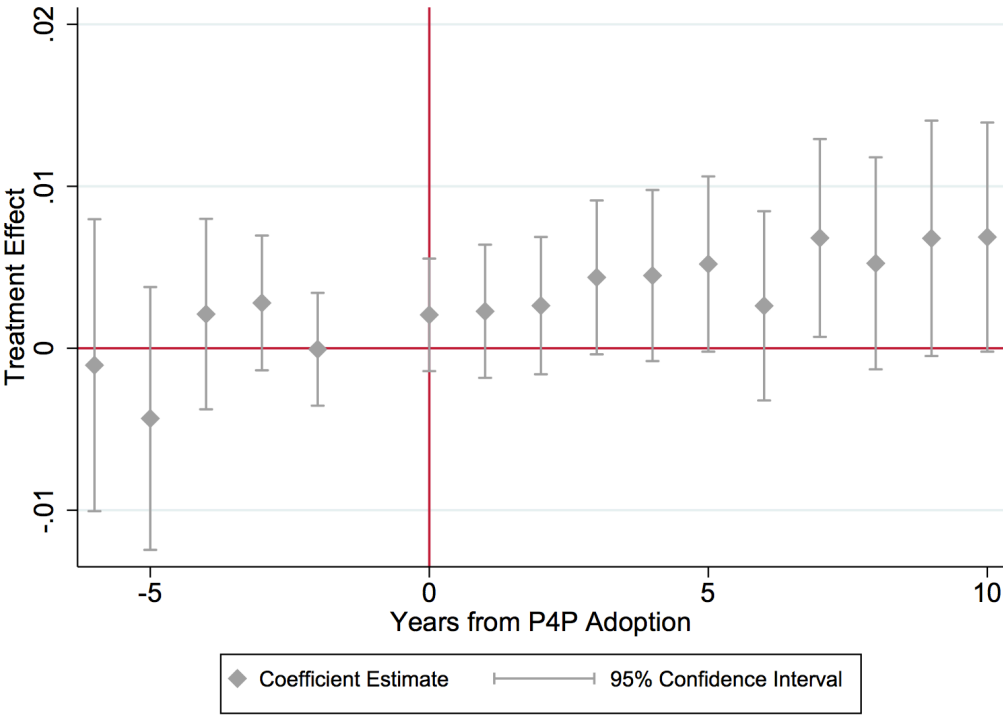
Note: Statistics Denmark's translation of the Danish language questionnaire.

Figure 2: Reported Pay-for-Performance Adoption Year in 1,309 Firms



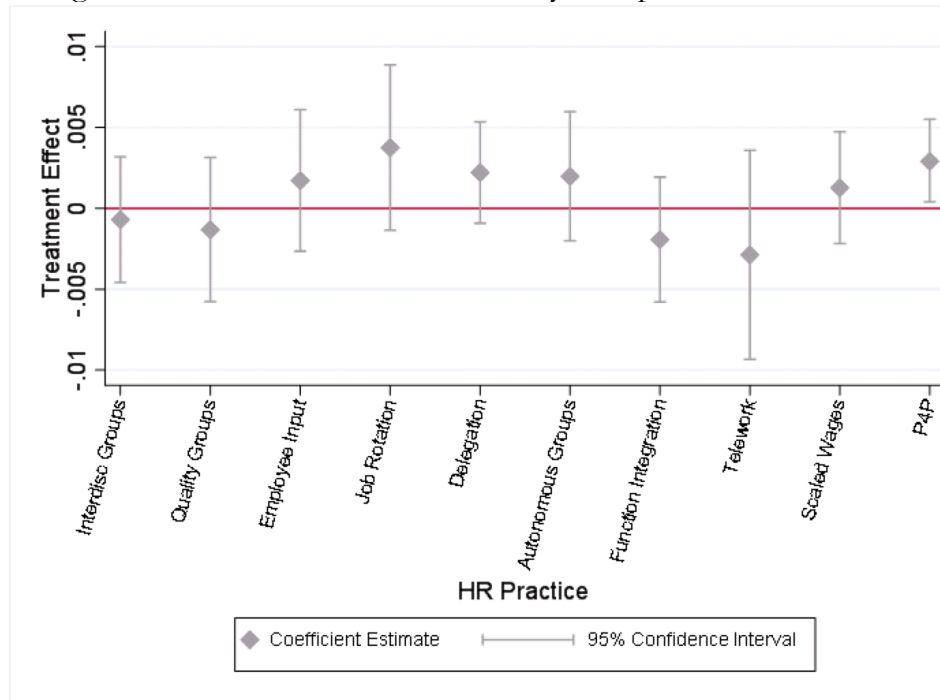
Notes: The vertical dashed lines represent the years of our mental health data. Our treatment effects are identified only off those firms adopting P4P within this time range.

Figure 3: Leads and Lags Estimates for Worker FE Model



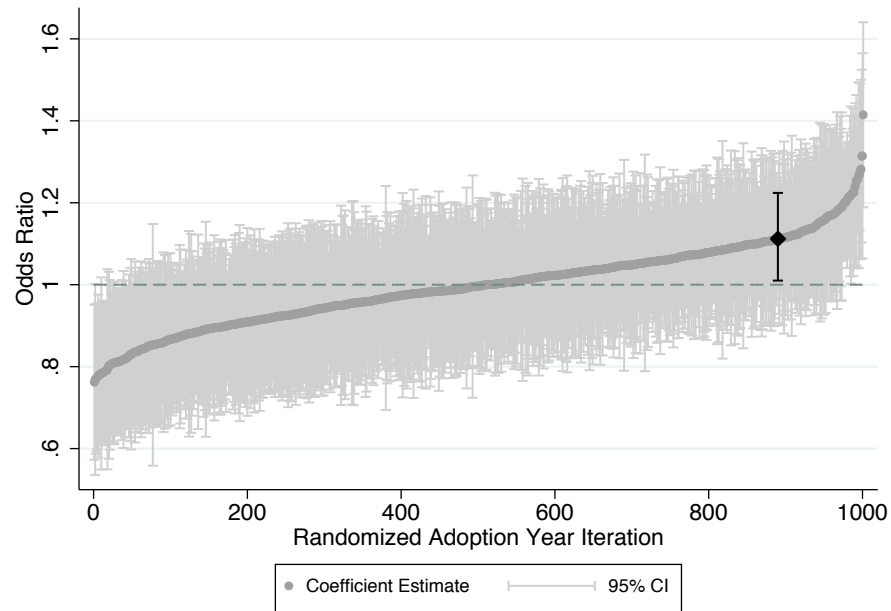
Notes: Each whisker plot represents the coefficient estimate and 95% confidence interval for the estimated treatment effect for each year before and after P4P implementation. The omitted year is the year before adoption.

Figure 4: Effect of Other HR Policy Adoptions on Mental Health



Notes: Each whisker plot represents the coefficient and confidence interval for one worker FE regression of *stress* on the adoption of a different HR policy.

Figure 5: 1,000 Placebo Tests of Individual FE Models



Notes: Each whisker plot represents the coefficient and confidence interval for one individual FE regression of *stress*, where each firm is randomly assigned a different firm's P4P adoption date. The larger and darker whisker represents the true data and coefficient present in Table 4, Column 1.

Figure 6: Theoretical Framework for Empirical Results

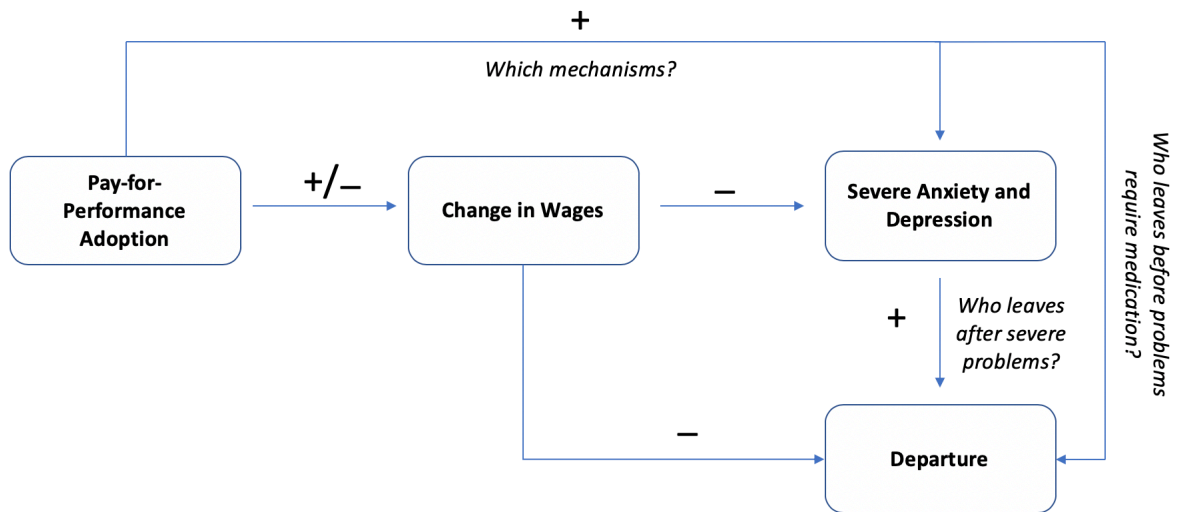


Table 1: Descriptive Statistics and Correlations for Employee-Year Data

[illegible]

Table 2: Pay-for-Performance Adoption and Mental Health Prescriptions

	(1)	(2)	(3)	(4)	(5)	(6)
Model:	OLS	OLS	OLS	OLS	OLS	OLS
Dependent Variable:	Stress	Benzo	SSRI	Stress	Benzo	SSRI
Post-P4P	0.0029** (.020)	0.0021* (.079)	0.0016* (.053)			
Ln (P4P Percentage)				0.0009** (0.020)	0.0006* (0.079)	0.0004 (0.116)
Year Controls	Yes	Yes	Yes	Yes	Yes	Yes
Family	Yes	Yes	Yes	Yes	Yes	Yes
Worker FE	Yes	Yes	Yes	Yes	Yes	Yes
Firm Employment Size	Yes	Yes	Yes	Yes	Yes	Yes
Adjusted R^2	0.482	0.447	0.477	0.632	0.609	0.627
Observations	1,159,417	1,159,417	1,159,417	1,092,608	1,092,608	1,092,608

Note: P-values presented in parentheses, and are calculated based on standard errors block-bootstrapped at the firm level using 500 reps. R-squared measures for OLS include fixed effects. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 3: Firm Fixed Effect Models

	(1)	(2)	(3)	(4)	(5)	(6)
Model:	OLS	OLS	OLS	OLS	OLS	OLS
Dependent Variable:	Stress	Benzo	SSRI	Stress	Benzo	SSRI
Post-P4P	0.0023** (0.032)	0.0011 (0.409)	0.0013* (0.066)			
Ln (P4P Percentage)				0.0008** (0.035)	0.0015 (0.163)	0.0003 (0.171)
Year Controls	Yes	Yes	Yes	Yes	Yes	Yes
Family	Yes	Yes	Yes	Yes	Yes	Yes
Demographics	Yes	Yes	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes
Firm Employment Size	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.026	0.025	0.001	0.025	0.025	0.009
Observations	1,159,417	1,159,417	1,159,417	1,092,608	1,092,608	1,092,608

Note: P-values presented in parentheses, and are calculated based on standard errors block-bootstrapped at the firm level using 500 reps. Reduced observations for columns (4)- (6) due to non-response for P4P percentage. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 4: Medication Changes by Gender (Worker Fixed Effects)

	(1)	(2)	(3)	(4)	(5)	(6)
Sample:	Women	Men	Women	Men	Women	Men
Model:	OLS	OLS	OLS	OLS	OLS	OLS
Dependent Variable:	Stress	Stress	Benzo	Benzo	SSRI	SSRI
Post-P4P	0.0033 (0.169)	0.0031* (0.062)	0.0009 (0.666)	0.0019 (0.135)	0.0018 (0.362)	0.0017* (0.064)
Year Controls	Yes	Yes	Yes	Yes	Yes	Yes
Family	Yes	Yes	Yes	Yes	Yes	Yes
Worker FE	Yes	Yes	Yes	Yes	Yes	Yes
Firm Employment Size	Yes	Yes	Yes	Yes	Yes	No
Adjusted R^2	0.481	0.469	0.442	0.442	0.499	0.445
Observations	404,341	755,076	404,341	755,076	404,341	755,076

Note: P-values presented in parentheses, and are calculated based on standard errors block bootstrapped at the firm level with 500 reps. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 5: Medication Changes by Gender (Firm Fixed Effects)

	(1)	(2)	(3)	(4)	(5)	(6)
Sample:	Women	Men	Women	Men	Women	Men
Model:	OLS	OLS	OLS	OLS	OLS	OLS
Dependent Variable:	Stress	Stress	Benzo	Benzo	SSRI	SSRI
Post-P4P	-0.0015 (0.614)	0.0045*** (0.000)	-0.0023 (0.380)	0.0030*** (0.009)	-0.0003 (0.865)	0.0021** (0.015)
Year Controls	Yes	Yes	Yes	Yes	Yes	Yes
Family	Yes	Yes	Yes	Yes	Yes	Yes
Demographics	Yes	Yes	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes
Firm Employment Size	Yes	Yes	Yes	Yes	Yes	No
R^2	0.035	0.019	0.038	0.019	0.013	0.008
Observations	404,341	755,076	404,341	755,076	404,341	755,076

Note: P-values presented in parentheses, and are calculated based on standard errors block bootstrapped at the firm level with 500 reps. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 6: Medication Changes by Age Group (Worker Fixed Effects)

	(1)	(2)	(3)	(4)
Sample:	Age 20-29	Age 30-39	Age 40-49	Age 50-65
Model:	OLS	OLS	OLS	OLS
Dependent Variable:	Stress	Stress	Stress	Stress
Post-P4P	-0.0018 (0.426)	0.0033 (0.106)	0.0005 (0.855)	0.0077*** (0.008)
Year Controls	Yes	Yes	Yes	Yes
Family	Yes	Yes	Yes	Yes
Worker FE	Yes	Yes	Yes	Yes
Firm Employment Size	Yes	Yes	Yes	Yes
Adjusted R^2	0.398	0.449	0.533	0.574
Observations	274,830	328,665	270,897	258,094

Note: P-values presented in parentheses, and are calculated based on standard errors block bootstrapped at the firm level with 500 reps. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 7: Medication Changes by Age Group (Firm Fixed Effects)

	(1)	(2)	(3)	(4)
Sample:	Age 20-29	Age 30-39	Age 40-49	Age 50-65
Model:	OLS	OLS	OLS	OLS
Dependent Variable:	Stress	Stress	Stress	Stress
Post-P4P	-0.0019 (0.245)	0.0008 (0.739)	-0.0005 (0.878)	0.0104** (0.018)
Year Controls	Yes	Yes	Yes	Yes
Family	Yes	Yes	Yes	Yes
Demographics	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes
Firm Employment Size	Yes	Yes	Yes	Yes
R^2	0.01	0.014	0.022	0.026
Observations	274,830	328,665	270,897	258,094

Note: P-values presented in parentheses, and are calculated based on standard errors block bootstrapped at the firm level with 500 reps. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 8: Medication Changes by Wage Changes

	(1)	(2)	(3)
Model:	OLS	OLS	OLS
Dependent Variable:	Stress	Stress	Stress
Post-P4P For Wage Losers	0.0039* (.077)		
Post-P4P For Wage Gainers	0.0004 (.862)		
Post-P4P For Below Average Wage Changes		0.0042* (.055)	0.0039** (.036)
Post-P4P For Above Average Wage Changes		0.0012 (.488)	0.0010 (.601)
Year Controls	Yes	Yes	Yes
Family	Yes	Yes	Yes
Worker FE	Yes	Yes	Yes
Firm Employment Size	Yes	Yes	Yes
Adjusted R^2	0.482	0.439	0.409
Observations	1,159,417	1,159,417	1,159,417

Note: Model 1 provides separate treatment effects for those who lose or gain income in the first year after P4P introduction. Model 2 separates workers by above- or below-mean wage change in these years. Model 3 uses below- or above-median wage change. P-values presented in parentheses, and are calculated based on standard errors block-bootstrapped at the firm level using 500 reps. R-squared measures for OLS include fixed effects. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 9: Turnover Based on Mental Health and Performance

Model:	(1)	(2)	(3)	(4)	(5)
Dependent Variable:	Leave	Leave	Leave	Leave	Leave
Medication Usage	0.062*** (.000)	0.057*** (.000)	0.086*** (.000)		
P4P Use at Firm	-0.017 (.121)	-0.017 (.125)	-0.017 (.128)	0.067*** (.000)	0.117*** (.000)
Medication Use X P4P Use	0.011* (.067)	0.011* (.052)	0.009 (.380)		
Low Wage Trajectory				Absorbed	Absorbed
Low Wage X P4P Use				0.024** (.0012)	0.039*** (.002)
Year Controls	Yes	Yes	Yes	Yes	Yes
Family	Yes	Yes	Yes	Yes	Yes
Worker FE	Yes	Yes	Yes	Yes	Yes
Firm Employment Size	Yes	Yes	Yes	Yes	Yes
Adjusted R^2	0.078	0.076	0.076	0.124	0.133
Observations	1,159,417	1,159,417	1,159,417	106,337	140,642

Note: Leave indicates the employee chose to switch jobs, enter unemployment, or retire early in that year. Model 1 uses both SSRI and benzodiazepines, while Model 2 uses only benzo and Model 3 only SSRI. Model 4 separates workers by above- or below-mean wage change in the two years prior to P4P adoption. Model 5 uses below- or above-median wage change. P-values presented in parentheses, and are calculated based on standard errors block-bootstrapped at the firm level using 500 reps. R-squared measures for OLS include fixed effects. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Michael S. Dahl is a Professor of Entrepreneurship and Organizational Behavior at the Department of Management, Aarhus University, Denmark. He has a PhD (2004) and a MSc in Industrial Economics (1999) both from Aalborg University. Previously, he held visiting scholarships at Carnegie Mellon University, SCANCOR at Stanford University, and Eindhoven Center for Innovation Studies.

Lamar Pierce is Professor of Organization & Strategy and Associate Dean at the Olin Business School at Washington University in St. Louis. At Olin, he teaches Strategic Management, Business, Government, and Society, and Incentives and Accountability. At the Brookings Institution, where he is a Non-Resident Senior Fellow, he teaches Ethics and Human Capital to civilian and military leaders at the Brookings Institution